Preprints.org

Article

Data Quality of Different Modes of Supervision in Classroom Surveys

Till Stefes [*]

Article

Data Quality of Different Modes of Supervision in Classroom Surveys

Till Stefes [1,2,*]

[1] Ruhr-University Bochum; till.stefes@rub.de

[2] FH Münster; ts752354@fh-muenster.de

[*] Correspondence: till.stefes@rub.de; Tel.: +49 234 32 22782
    ORCID-ID: 0000-0002-5877-6114

Abstract: Conducting quantitative research involving adolescents demands a thoughtful approach to the question of supervision, given that each option comes with its distinct set of implications. This study reviews these implications and empirically tests whether differences in data quality can be found among three modes of standardized survey research with medium sized groups of adolescents (12-17 years). Data basis is a quasi-experimental survey study testing different forms of digital, hybrid or in-person supervision, that took place in 2021 in secondary schools in Germany (N=923). Aim of this study is testing how aspects of data quality – item-nonresponse, interview duration, drop-out-rate and response patterns – differ between these forms of supervision. Results could help researchers surveying young people to decide (1) whether they allow confidants or other adults to be present during interviews, (2) if they can rely on teachers alone when surveying classrooms and (3) if it is cost-efficient to send out external supervisors for classroom sessions. While dropout-rates do not differ, item non-response, interview duration, and response patterns differ significantly, students supervised at home by external interviewers answered more questions, took more time to answer and are less likely to give potentially meaningless answers in grid questions. The implications drawn from the findings question the common approach to solely rely on teachers for survey administration without the support of external supervisors or adequate training. Recruiting respondents via schools and surveying them online in their homes during school hours shows to be a robust method with regard to the analyzed indicators.

Keywords: Standardized survey research, nonresponse, adolescents, data quality, third party effects, interviewer effects, school, education

1. Introduction

Reviewing survey errors is essential for maintaining the accuracy of self-administered questionnaire data. This process helps identify and minimize errors like sampling, nonresponse, and measurement issues specific to self-administered questionnaires (Ziegel 1990). By addressing these errors, researchers can enhance the reliability and validity of future survey results.

While ensuring data quality is essential in any research project, it becomes especially challenging when surveying adolescents. They are a heterogeneous group, with some being forthcoming and others being rebellious or unwilling to answer personal questions. In addition, adolescents tend to be impatient and restless, especially when surveyed in groups. These factors can complicate the survey process and potentially compromise data quality and the independence of individual responses. Thus, adequate supervision might identify and mitigate any sources of bias that may arise in order to ensure accurate and reliable results.

Schools are key institutions in adolescent life and an ideal setting for research dealing with the topic (Hatch et al. 2023). A typical approach to survey adolescents is conducting standardized surveys in classrooms with teachers or external supervisors ensuring consistency of the process (Hallfors et al. 2000; Lucia et al. 2007; Alibali & Nathan 2010, Bartlett et al. 2017, March et al. 2022). Classrooms are especially convenient for several reasons. Working with schools is time-saving because we can

interview whole classes at once and don't have to visit homes (Walser & Killias 2012) or wait for postal answers. Surveys conducted in schools are considered more valid compared to surveys taking place at home (Kann et al. 2002). Brener et al. (2006), for instance, found a 25% lower likelihood of reporting offending behaviour when collecting data in homes rather than schools. Cops et al. (2016) also found significantly higher prevalence of delinquency when youths were filling out questionnaires in school versus a mail survey design. Some topics or target groups are difficult to reach, recruitment via public schools is an efficient strategy to reach representative populations (Ellonen et al. 2023). Others point out that research in schools add the risk of selection bias on the school level, depending on which schools are selected for the sample (Newransky et al. 2020).

Given successful recruitment of schools and classrooms, teachers can be pivotal for a survey's success: parents trust them (parental consent, unit nonresponse), students rely on their expertise (item nonresponse, validity) and are expected to respect their authority (unit & item nonresponse, validity). However, relying on them to convey a scientific survey might bias results, because of (possibly even unnoticed) influence on the students' response behaviour (see Strange et al. 2003, Heath et al. 2013; Rasberry et al. 2020). Many questionnaires contain questions about delinquency, drug abuse, school, relationships to adults in school and further topics that are prone to social desirability bias. The presence of external supervisors may potentially have an impact, as suggested by Bidonde et al. (2023); they might emphasize the significance of maintaining neutrality among the adults present, either through verbal communication or simply by being physically present.

Many studies examining with effects of the presence of third parties focus on face-to-face or telephone interviews (e.g. Deakin & Wakefield 2014, Weller 2017, Hennessey et al. 2022, Goh & Binte Rafie 2023). Much research on self-administered survey research is evaluating methods of (mental) health of adolescents (e.g. Kann et al. 2002; Brener et al. 2006; Raat et al. 2007; Rasberry et al. 2020; Bidonde et al. 2023; Ellonen et al. 2023; Hatch et al.) or criminology and delinquency of young adults (e.g. Lucia et al. 2007; Walser & Killias 2012; Kivivuori et al. 2013; Cops et al. 2016; Gomes et al. 2019). However, meta-studies still demand more systematic evaluation of the effects of supervision on response behaviour (Gomes et al. 2019; Bidonde et al. 2023).

Although there is extensive research dealing with mode effects, additional research with controlled designs is necessary to find effective methods and strategies that ensure satisfactory response rates among adolescents in mental health surveys (Bidonde et al. 2023) and to find out how different settings might affect response behaviour (Gomes et al. 2019).

Given the sensitive environment of schools and the difficult task to get qualitatively good data from minors, social science must constantly evaluate its methods (Newransky et al. 2020; Hatch et al 2023). Plus, the many challenges of survey research in this setting and frequently low research budgets call for justified decisions on how to carry out the aspired survey (Walser & Killias 2012). Two questions arise: is teacher supervision a threat to data quality, and is investment external supervision necessary?

To examine effects of different modes of supervision during standardized survey research on data quality, this publication utilizes dataset resulting from a quasi-experimental study design. The survey study under investigation is based in Germany's largest metropolitan area, testing different forms of digital, hybrid and in-person supervision. *UWE* („Umwelt, Wohlbefinden und Entwicklung" = "Environment, Well-Being and Development") is a classroom-based, repeated cross-sectional study. It serves multiple purposes, among them sociological research of adolescent life, but more importantly it shall empower youths to have their voices heard by school and municipality officials. Aim of this publication is testing how item-nonresponse, interview duration, drop-out-rate and response patterns differ between groups, depending on their supervision. Results could help researchers surveying youths and adolescents to decide whether they allow confidants or other adults to be present during (group-) interviews, if they can rely on teachers alone when surveying classrooms, and if it is cost-efficient to send out external supervisors for classroom sessions.

1.1. Previous Research

Researchers surveying young people during the pandemic have faced similar challenges all over the world (e.g., Gassman-Pines et al. 2020; Goh & Binte Rafie 2023) and have found similar workarounds, although most of them had to start from scratch. Until very recently, there has not been much published research on the effects of lockdown-workarounds on survey data quality. As the term "workaround" suggests, much of the conducted research during that time, the present study included, used ad-hoc methods, fitting individual needs and available infrastructure. Assuming the researchers in the field had profound knowledge of survey methodology and put it to good use when developing work arounds, scholars can now benefit from the resulting pioneer work.

The following section will present previous research on third party effects, effects on external supervision and relevant literature on survey research in general, that is relevant when surveying youths and adolescents with standardized questionnaires in classrooms, using different modes of supervision.

1.2. Supervision during Self-Administered Surveys

In theory, a well-designed questionnaire should not require any supervision at all, but evidence on data quality is not unambiguous. For instance, Felderer et al. (2019) found that web surveys tend to have higher nonresponse rates than surveys led by interviewers. Their results contradict the findings of Mühlböck et al. (2017), who investigated self-completion surveys among young adults and found no significant differences in response behaviour between web surveys without supervision and modes with interviewers present – although completion rates were higher in the latter group. However, self-administration has been shown to be less prone to response bias (Felderer et al. 2019). Atkeson et al. (2014) argue that the presence of an interviewer can alter response patterns on questions that pertain to an individual's personal beliefs, attitudes, or experiences. Bidonde et al. (2023) found that response rates of adolescents in studies of mental health can indeed vary with survey mode, consent type, and incentives used. Overall, it seemed that when there is any kind of supervision involved, response rates were at least slightly higher. The results of Cops et al. (2016) suggest that the mode of administration can impact response behaviour on different levels. Survey mode influence extends to both individuals' likelihood to participate in the study, eliciting selection bias, and the potential for differential tendencies to report criminal behaviour among participating individuals, prompting measurement bias. Thus, variations in the prevalence estimates of criminal behaviour between studies arise from differences in the participating population, as well as potential effects related to the setting or anonymity. Since young people are considered less likely to participate in surveys compared to older people (Ziegel 1990), a closer look into potentially enhancing methods is worthwhile.

1.3. Effects of Teachers' Presence

Recruiting and surveying respondents in schools is a rather inexpensive and an efficient way to achieve high response rates (Heath et al. 2009: 146, Alibali & Nathan 2010, Bartlett et al. 2017, March et al. 2022) and representative samples (Ellonen et al. 2023). An important prerequisite for successful research involving schools and school personnel is to tailor survey research designs to schools' needs (Hatch et al. 2023). However, the study presented by Rasberry et al. (2020) shows how difficult survey research in schools can be. Their study is prepared and conducted carefully, and every pitfall seems to be anticipated, in the field they still struggled with teachers failing to monitor the survey or unprecedented issues with post-survey logistics.

When surveys take place in schools, it is common practise to work with teachers as "assistant supervisors", as they are already figures of authority in the target groups and can usually handle group dynamics and disruptions. Although their ability to do so may vary (March et al. 2022), this should increase data quality in terms of validity and completion rates (Bidonde et al. 2023).

A typical notion in the literature is that teachers are important or at least very helpful to get the survey infrastructure setup (e.g., in Strange et al. 2003; Heath et al. 2009). They can help keeping respondents' motivation up and help retrieve basic information, such as zip codes. These are typically used to obtain small-scale data necessary to localise neighbourhoods that require municipal attention.

Retrospective data might be more accurate, when a confidant can help remembering things. Teachers can provide valuable information and insights about academic abilities and behaviour. All the above is also noted in the protocols of the *UWE* survey. In conclusion, there are arguments speaking in favour of teachers facilitating the survey process:

Hypothesis 1a: When conducting standardized survey research in classrooms, the presence of teachers increases data quality.

When conducting or assisting with a survey in their classroom, teachers may find themselves in a unique position where they need to balance two roles. They act as supervisors but are a third party at the same time. Yet, none of these attributions fits perfectly and they are also ambiguous in their potential effects. In their role as supervisors, they are handing out questionnaires or online survey links and are the primary source of information on comprehension questions. In many cases, they also read an introduction. From the researcher's perspective however, they are considered as a third party. Primarily because they have a personal relationship to the respondents jeopardizing their neutrality. In addition, they usually don't know the questionnaire at all. In Demkowicz et al. (2020), students reported that their teachers were not very helpful when asked comprehension questions. Before the survey takes place in their classroom, teachers have not been trained as interviewers in most cases, although Hatch et al. (2023) highly recommend that. Depending on their seniority, they might or might not have been part of research in school (March et al. 2022). Hence, it cannot be assumed that they are aware of how their assistance can affect the accuracy of responses. Rasberry et al. (2020) even report teachers filling out the survey themselves, discussing survey questions or reading sensitive questions out loud, potentially preventing students from answering truthfully.

The literature knows further sources of potential influence teachers or other adults can have on young people's response behaviour. An obvious one is social desirability bias. Social desirability bias is the tendency of respondents to provide answers that are socially acceptable rather than their true opinions or behaviours, which is more likely to occur in interviewer-administered surveys (Atkeson et al 2014). Adolescents may be particularly prone to social desirability bias because they are often highly influenced by their peers and social norms (Ziegel 1990; Brown & Larson 2009). Cops et al. (2016) found that self-reported delinquency was significantly lower when youths were supervised by adults close to them, which implies that third parties should be avoided and surveys with sensitive questions should rather not take place at home. As Tourangeau and Smith (1996) found in an early review of studies using the computer assisted personal interview (CAPI) approach, social desirability bias can be reduced by using self-administered questionnaires or computer-assisted self-interviewing, which can increase anonymity and decrease social pressure to conform.

In contrast, the mere presence of teachers might increase the latter, since they are authority figures (Möhring & Schlütz 2019, 49). They establish the typical classroom-atmosphere and thus, respondents likely find themselves in their social role as students. When asked how they feel in this role, the presence of involved teaching personnel likely influences their response behaviour. The relationship between students and teachers could also increase item non-response: Strange et al. (2003) report that students were hesitant when asked about personal information by their own teachers.

Duncan and Magnuson (2013) discuss potential bias arising from the presence of adults during surveys in school settings due to differences in socioeconomic status. This might be especially relevant in this particular case: Teachers in Germany are highly qualified personnel and thus have a high socio-economic status. The survey analysed in this study was conducted in two of the poorest cities in Germany, so there is a high share of youths from materially deprived households.

The difference between a survey and an exam can be difficult to internalize for both students and teachers. A frequent observation from the protocols of *UWE* is that teachers tend to rush things. Of course we want respondents to take their time, teachers however are used to tight schedules and like to get things done in time. This may partly stem from the practical necessity for schools to efficiently manage their staff resources, given that they are frequently understaffed, leading them to

be cautious about allocating excessive time for survey projects (March et al. 2022). Moreover, surveys are often perceived as additional work by teachers, which is not wrong (ibid.). Alibali and Nathan (2010) as well as Hatch et al. (2023) strongly recommend being prepared for that and being patient with schools and teaching staff as their time is limited.

The problem is recognized and not easily surmountable, so we need to assess whether it has adverse consequences. Strange et al. (2003) showed that when strict time limits are established, those with literacy problems show higher drop-out rates, and consequently students from lower social classes were less likely to complete their survey. Gummer and Roßmann (2015) found that longer interview duration is related to higher motivation among respondents – strict time limits might dampen this motivation or at least eradicate its positive impact on data quality (e.g. item-nonresponse). Following these arguments against teachers' presence, a counterhypothesis challenging my first one would be:

Hypothesis 1b: When conducting standardized survey research in classrooms, the presence of teachers reduces data quality.

1.4. Effects of External Supervision

The obvious solution to this dilemma would be to educate teachers on the science of survey research, like Hatch et al. (2023) did. Unfortunately, this can be very difficult to pull off for various reasons, with time being the most pressing one. A potential remedy is the presence of a third, neutral party that is familiar with the pitfalls of conducting surveys. Trained interviewers can intervene when teachers are unintentionally biasing responses and help with comprehension questions. Communication between members of the research team can also help teachers and respondents to understand the purpose of the study, resulting in higher motivation among all parties involved (March et al. 2022).

Demkowicz et al. (2020) report that respondents were motivated to complete their questionnaire because they were aware that their participation might be helpful to others in the future. Reflecting on their own lives and well-being has been seen to be helpful to children and young people as well, again increasing motivation to complete the questionnaire. Those who are actively engaged in the projects associated with the survey (e.g. researchers acting as supervisors or even hired, but trained staff) are more likely to effectively convey to respondents the significant impact their participation can have on the project's success and its subsequent benefits for youths and adolescents.

Hypothesis 2a: When conducting standardized survey research in classrooms, the presence of external supervisors increases data quality.

According to Epstein (1998), adults who work in a school environment can be perceived as "another teacher" by both students and teachers. This can have an impact on the objectivity of interviewers and may also influence how young people respond. Or if it doesn't, sending in supervisors might not even worth the effort, which cannot be underestimated (see Walser and Killias 2012, Bartlett et al. 2017; March et al. 2022). Strange et al. (2003) found that when surveys were administered to students either by a teacher or a researcher, there was no significant difference in the likelihood of students completing the questionnaire. Walser and Killias (2012) and Kivivuori et al. (2013) dealt with highly sensitive questionnaires and also found little significant differences in response behaviour of juveniles supervised by teachers versus external supervisors. The opposing hypothesis must therefore be:

Hypothesis 2b: When conducting standardized survey research in classrooms, the presence of external supervisors does not increase data quality.

Demkowicz et al. (2020) discuss the role of non-teaching staff in creating an environment that deviates from the typical classroom atmosphere and empowers respondents with a choice regarding their participation. Ethically, it is highly preferable for respondents to be aware that survey participation is voluntary. Moreover, informed consent can significantly enhance completion rates (ibid.). However, participation rates can also benefit from what Demkowicz et al. refer to as a 'fait accompli' scenario (2020: 11). This refers to situations where parents and teachers of the respondents have already agreed to the surveys, and the assigned schoolwork is generally compulsory. Gaining consent from school and parents in the first place is a field of research in itself (see Alibali & Nathan 2010; Bartlett et al, 2017; Hatch et al. 2023)

1.5. Effects of Using Video Conference Software

Cost efficiency has been the primary challenge associated with deploying external supervisors or trained interviewers for surveys until recently. Professional interviewers are expensive and survey projects usually run on a budget (Walser & Killias 2012). In 2020 and following, another big problem superseded that issue: contact limitations and closing of schools. With contact limitations and lockdowns in place, sending in external supervisors was literally impossible.

One solution for that problem was the increasingly common use of video conference software, which allows people to virtually communicate face-to-face without having to be in the same room. Some schools were already using it for distance learning and as many others, *UWE* took advantage of their efforts. A complex coordinating process generated a set of new survey modes.

Video conference software is already used quite regularly in qualitative social science (see e.g. Deakin & Wakefield 2014, Weller 2017, Hennessey et al. 2022, Goh & Binte Rafie 2023). Common sources of bias are along the same lines as in face-to-face or telephone interviews (ibid.). An additional one is representativeness, as not all social strata have equal access to and competence to use VCT.

For quantitative research and questionnaire-based surveys it is uncommon to make use of it, because it may appear impractical or redundant. Impractical, because it takes effort to arrange and setup despite being technically unnecessary. Redundant, because a good questionnaire speaks for itself and does not need an interviewer. As discussed above, when dealing with adolescents in larger groups, supervision is required. There tend to be disruptions and group-dynamics that are unique for this age group, and that anyone can imagine who has been working with teenagers before. Strange et al. (2003) described them vividly. Simply put, we cannot assume independent observations in shared classrooms. Finally, if external supervision is needed, VCT offers a relatively inexpensive solution.

1.6. Web-Surveys

The age-group analysed here is born after 2004, individuals belonging to the generation we consider "digital natives". It can be argued that their extensive experience with digital devices would not result in any significant differences in data quality when completing a questionnaire with a pen, a school-PC, or a smartphone. Raat et al. (2007) found negligible differences between adolescents answering a health-related survey either on paper in schools or via web survey in feasibility, reliability, and validity of the scales. Hallfors et al. (2000) report that self-administered surveys decrease item nonresponse compared to paper forms.

Young respondents in this generation, born after 2000, might even prefer the digital delivery. In the study conducted by Demkowicz et al. (2020), respondents expressed a preference for digital delivery, citing reasons such as increased efficiency, their familiarity with surveys on digital devices, concerns about anonymity due to recognizable handwriting, and heightened security concerns associated with the potential loss of paper forms.

Gummer and Roßmann (2015) argue convincingly, that the device matters indeed. However, their main argument was that the questionnaire design must fit the device, because there may be differences in visibility of the questionnaire or different download-times when using mobile devices – affecting response latency. Few respondents in Demkowicz et al. (2020) did also state that they

struggled sometimes with the visual formatting. The survey project analysed here utilized software that is suitable for all devices ("Sosci-Survey" (Leiner 2021)) and tested the resulting questionnaire on all possible devices – there was satisfactory visibility of the questionnaire and there were no reports about problems with downloading the questionnaire or uploading responses.

2. Materials and Methods

The hypotheses will be tested in the following using regression analyses, based on data from the 2021 wave of the survey study *UWE* that has been described briefly above (available via Stefes 2023). Data quality indicators will be predicted based on different supervision modes, while controlling for respondent- and interview characteristics.

UWE set out asking every youth in grades seven and nine in two Ruhr-Area municipalities about their well-being, everyday life, and social resources, every other year since 2019 using standardised questionnaires (see Schwab et al. 2021 and Stefes et al. 2023). The analyses in this study utilize data resulting from the survey round in one municipality that enabled cooperation with local secondary schools in spring 2021. Within this wave, a cohort comprising 923 students from grades seven and nine, typically ranging from 12 to 15 years old, within a single municipality, has commenced the process of responding to the questionnaires. During this period, schools were closed or opened for a limited number of students, depending on incidence rates of the Covid-19-pandemic. Due to these unique circumstances, different modes of supervision were used in the same schools and even classrooms were surveyed in various setups.

The survey itself covers multidimensional operationalisation of subjective well-being, social resources and contexts and allows drawing a comprehensive picture of adolescent life from a socio-ecological perspective (Knüttel et al. 2021). Initially, the data collection involved handing out questionnaires in classrooms with an external supervisor and a teacher present. Supervisors were responsible for explaining and answering questions, while teachers represent figures of trust and authority during these sessions. The supervisors were mostly the researchers responsible for the survey itself and their student assistants, who have been trained as interviewers.

The study provided three modes of supervision for groups of respondents aged 12 to 15, including (A) only teachers present, (B) teachers and external supervisors present via video-conference technology (VCT), or (C) external supervisors present via VCT without teachers. Questionnaires were self-completed using school-owned devices, personal devices, or paper forms. This quasi-experimental study allows for systematic evaluation of potential differences in data quality-between these modes of supervision.

2.1. Data Quality indicators and Analyses

There are three factors of data quality that can be indicated reliably using the data sets of *UWE*: Dropouts, item nonresponse, interview duration and straightlining. This section describes these indicators and discussed why they are used and how they are analyzed.

A response is defined as drop-out, when the survey ended before the last quarter of the questionnaire was answered. The drop-out threshold lies between two item batteries, more general questions about school life and a battery about bullying. Roughly 5% of all respondents answered less than 75% of the overall questionnaire, this is what is considered dropping out prematurely. The indicator can be seen as a negative of completion rates, which is a common measure of data quality. Mühlbrock et al. 2017 used drop-out rates to compare self-administered versus supervised surveys among young adults but found no significant differences in drop-out rates. Drop-out is a binary variable, hence the use of logistic regression is adequate to predict it (Heeringa et al. 2017, Niu 2020). Logistic regression assumes a linear relationship between the independent variables and the log-odds of the dependent variable. It also assumes the absence of multicollinearity, meaning that independent variables are not highly correlated. Multicollinearity concerns have been effectively addressed, as evidenced by satisfactory Variance Inflation Factor (VIF) values within the regression model, affirming the absence of significant multicollinearity among the independent variables. Limitations include its sensitivity to outliers and the assumption of linearity, which may not always hold in

complex relationships. Additionally, logistic regression assumes that the observations are independent, and violations of this assumption can affect the accuracy of parameter estimates (Niu 2020). The independence of observations assumption usually does not hold in classrooms, as all respondents in the same classroom are exposed to the exact same conditions, which can differ between classrooms. Therefore, clustered standard-errors and controls of all available conditions are implemented in the logistic regression model.

As a second indicator, I analyse item-nonresponse. There were 210 items presented to all respondents. The great number of items is one of the reasons why the absence of supervision was not considered in the field. Most respondents answered most of the questions, but there are also a lot of gaps in the data. While nonresponse rate is a common indicator for survey data quality (e.g. Gummer et al. 2021; Leiner 2019), its use as a proxy indicator for nonresponse bias has been questioned (Wagner 2010). Wagner (2010; 2012) calls researchers to use the fraction of missing information instead. They conclude that although completion rates might not be a good indicator for nonresponse bias, they are adequate for a comparison of different data collection methods (Wagner 2012). For the analyses in this study, item nonresponse is measured by simply counting the items with missing values in the data set. Some questions were filtered and could not be answered by all respondents, such as follow-up questions about migration background. They are excluded from the count. Respondents who were defined as drop-outs have been excluded form this analysis, because they are extreme outliers. The decision against a share or missing information instead of the raw count is based on the argument that a count is easier to interpret than a share, and the transformation into a percentage adds no information. Since the resulting variable can be described as count data, a variation of poisson regression is adequate (Little 1988; Little 1992). There is no reason to assume zero-inflation, because all zeros in the data are simply complete questionnaires. Therefore, a negative binomial model to predict the number of unanswered questions will be applied. This model, akin to the previously discussed logistic model, assumes independence of observations, which is addressed by using clustered standard errors. VIF-values cannot be estimated in negative binomial regression models. As recommended by Türkan and Özel (2013), the model utilizes jackknifed estimators to remedy potential effects of multicollinearity and reduce bias in the estimation process.

The duration of interviews conducted with digital questionnaire forms has been recorded, allowing examination for any indication of haste. Interview duration is an ambiguous indicator. We cannot simply claim that longer duration indicates higher quality than shorter, just as we assume less dropouts or nonresponses to be indicators of better quality. Questions can be answered too slowly or too quickly. The former might imply that respondents have trouble understanding the questions or being distracted. The latter could indicate speeding through the questionnaire or straightlining and consequently not answering truthfully. A very fast completion of the questionnaire can be a hint for a careless or fake response, hence Leiner (2019) suggests using completion times to identify meaningless data. In terms of response time for single items, Revilla and Ochoa (2015) mention that highly skilled respondent could answer very quickly, although extremely short response times rule out the possibility that respondents read the question at all. They seem to lean towards the notion that short response time is generally related to lower quality of responses. Their results support this claim convincingly and are in line with the findings of Gummer and Roßmann (2015), who found longer duration in higher motivated respondents. Interview duration was recorded in seconds during the self-administered questionnaire procedure. To make my results more comprehensible I recoded them into minutes. The analysis excludes dropouts and respondents that filled out paper forms. The former because their duration is irrelevant for this question, the latter because the required data could not be collected for this group. Since Tourangeau et al. (2013) used Cox regression to model interview duration, the study follows their recommendation here. Critical assumptions and limitations of cox regression are similar to the aforementioned and involve linearity of effects and independence of observations (Braekers & Veraverbeke 2005, Tourangeau et al. 2013). To assess potential inaccuracies due to multicollinearity, Variance Inflation Factors have been calculated and are satisfactory for all covariates.

Whether a response is truthfully or not is difficult to assess. Satisficing response behavior may lead respondents to anchor their answers on the first response option they find satisfactory, which easily minimizes the required effort for survey completion (Gummer et al. 2021). If they then align their subsequent responses with the initial choice, straightlining occurs. The absence of straightlining enhances the reliability of data by signaling genuine respondent engagement and truthful reporting (Leiner 2019; Gummer et al. 2021). *UWE* uses several item battery or grid questions suitable for a thorough examination of possible straightlining patterns. Many grid questions only contain three or four items, which can be answered truthfully with the same option repeatedly. Hence, the analysis presented in this work seeks patterns in item batteries containing at least five items. The questionnaire contains six grids that are unlikely to be answered truthfully by repeatedly choosing the same option. All of the items are answered on a five-point-scale. Another logistic regression model will predict how likely it is that at least one occurrence of straightlining can be detected in each supervision mode. The same measures are applied as in the first logistic model and all VIF-values are satisfactory.

2.2. Explanatory Variables

All statistical models control for the same characteristics of the respondents and the interview setting. Respondent characteristics include age, gender, migration background, German literacy, grade, and school type. School type distinguishes three types: higher secondary track (Gymnasium), comprehensive secondary track (Sekundarschule and Gesamtschule) and practical secondary track (Realschule and Hauptschule). The main difference is that attainment of the first one qualifies graduates for tertiary education, in the second this is optional and the third one does not offer this option. Table 1 presents descriptive statistics of the variables used in the analyses:

Table 1. Descriptive Statistics.

Variable	Mean	sd	Min	Max	N
Individual Characteristics					
Age	13.84	1.24	12	17	923
Gender: Female	0.54	0.50	0	1	923
Literacy Problems: Yes	0.02	0.15	0	1	923
Migration Backround: Yes	0.41	0.49	0	1	923
School: Comprehensive Secondary Track	0.38	0.49	0	1	923
School: High secondary Track	0.49	0.50	0	1	923
School: Practical Secondary Track	0.26	0.44	0	1	923
Interview Characteristics					
Delivery: Web survey	0.22	0.41	0	1	923
Teacher only (A)	0.49	0.50	0	1	923
Teacher & supervisor (B)	0.38	0.49	0	1	923
Data Quality Indicators					
Dropout: Yes	0.04	0.19	0	1	923
Item-nonresponse count	8.12	24.06	0	186	923
Interview Duration (Minutes)	21.03	10.76	2	56	923
Occurences of straightlining	0.28	0.66	0	4	923

Girls are slightly overrepresented in the sample. The maximum age in the sample is 17, although the usual age in grades seven to nine is 12 to 15. A very small portion of the sample has literacy problems, indicated by the question "How easy is it for you to read German?", to which they answered "hard" or "very hard". 41% of the sample has a migration background, which was defined as being born in another country or having a parent being born in another country than Germany. This share is consistent with the overall population of that age in metropolitan Germany.

The interview setting is characterised by the supervision (teacher present or not, supervisor present or not), the survey location (School or at home) and the form in which the questionnaire was delivered (paper or online). There were 84 classroom-sessions and the models control for potential

heteroskedasticity by clustering the standard-errors accordingly in all models. The supervision has three possible forms, assignment was done according to schools' equipment, teachers' preferences and the official contact-restrictions. An overview of the supervision modes used in which schools can be found in the Table 2. During the time the survey was conducted, some schools were practicing what was called "Wechselunterricht", or "alternate distance learning". Classes were split up in two groups, which were taught at home or in school for one week, in the next week they swapped places. This allowed the researchers to test different approaches on the same classroom, assignment to one of the two groups was usually random by surname.

(A) Teacher only: One share of classrooms was surveyed without external supervision involved. Teachers were given paper questionnaires or shortened online-survey links. The links were available as QR-codes as well. They had the option to present a video or explain the procedure themselves. For that option they had an introduction and manual prepared by the institute, in order to ensure that all respondents had the same information. Respondents could either use their personal devices, or school-owned devices, depending on teachers' perception what would work best in their classroom.

(B) Teacher and external supervisor: Another group of classrooms had teachers present and was more or less constantly connected to an external supervisor via video conference software. They introduced the survey and offered help with comprehension questions. They also observed the situation but could not effectively intervene, given their mere virtual presence. This mode was combined with online-surveys conducted on school-computers or respondents' personal devices.

(C) Supervisor only: Some classes were not available for being interviewed in school. We used the distance-learning channel they were already used to by that time to establish the group survey (Microsoft Teams in most cases). They were all online in a video call and filled out an online survey on their own devices at home. No teachers were present in this mode.

Table 2. presents the distribution of supervision modes among the three school-types and their response rates.

School	(A) Teacher only	(B) Teacher & supervisor	(C) Supervisor only	Response rate
Comprehensive secondary track 1				11%
Comprehensive secondary track 2				38%
Comprehensive secondary track 3				41%
High secondary track 1				49%
High secondary track 2				77%
High secondary track 3				64%
Practical secondary track 1				64%
Practical secondary track 2				53%
Practical secondary track 3				37%

Supervision Modes in surveyed schools.

3. Results

In the following predicted values for the indicators of data quality and full regression outputs are presented by mode of supervision.

3.1. Interview Drop-Outs

A drop-out can have several reasons. Participation is not mandatory. Questions are in part very personal and there are many of them. Presumably, respondents lost interest in the survey or felt that it was too personal. Another possible reason is that they could not finish in time. This affects item nonresponse and interview duration as well - the survey was conducted during school-hours (45 minutes), which puts a natural limit to interview-duration. The protocols mention that teaching personnel often insisted on ending the interview sessions when the "bell rang" (In most German

schools an emulated bell ring indicates the end of the lesson). Figure 1 reports the predicted probability of early interview drop out. Based on the logistic regression model (Table 3), the probabilities do not significantly differ and are below 0.8% - all three interview situations can offer an acceptable completion rate. There is no clear evidence regarding the hypotheses 1a and 1b. It might be that advantages and disadvantages of teachers supervising a survey cancel each other out. Hypothesis 2b stated that external supervisors do not matter for data quality and the result presented in Figure 1 hints in this general direction.
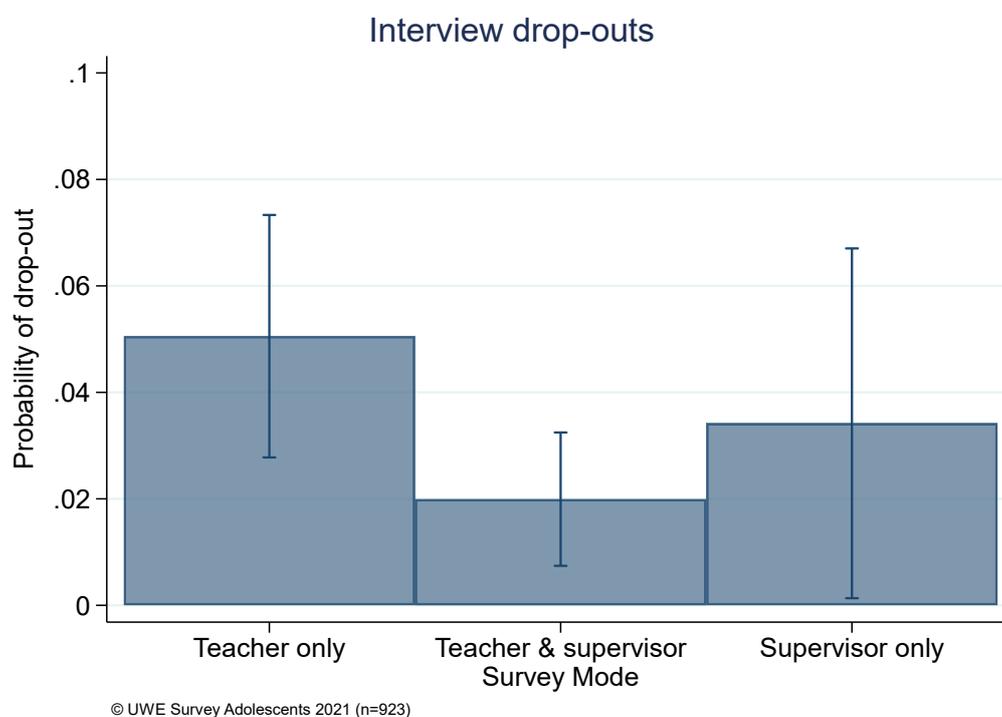

Figure 1. Predicted probability of dropout by mode of supervision (logistic regression, full figures in Table 2).

Table 3 presents the full model. A significantly higher likelihood of not finishing the survey was found in youths who have problems with reading in German. The coefficient indicates that falling in that category increases the log-odds of dropping out by 1.89. The fact that the coefficient showing a negative effect of supervision by both teachers and external supervisors, while the marginal effect is not significant in Figure 1 is likely due to interaction effects that are not considered in the model. Respondents that were not using paper forms did leave before finishing much more often. Students on the practical secondary track were much less likely to drop out than high secondary students. They were more often supervised by teachers only using paper forms, which might be a driving force here. Bias due to multicollinearity can be ruled out mostly, as the VIF-value of both variables are below 10.

Table 3. Interview drop-outs - logistic regression results.

	Coef.	St.Err.	z-value	p-value	[95% Conf	Interval]	Sig
Respondent characteristics:							
Gender: Female	0.069	0.341	0.20	0.84	-0.6	0.737	
Age	-.014	0.153	-0.92	0.36	-0.439	0.16	
Problems with reading German	1.829	0.778	2.35	0.019	0.305	3.353	**
Migration background	0.546	0.339	1.61	0.108	-0.119	1.21	

Interview characteristics:

Web survey	14.252	1.288	11.07	0	11.728	16.776	***
Teacher only (base)	-	-	-	-	-	-	
Teacher & supervisor	-1.331	0.487	-2.73	0.006	-2.286	-0.376	***
Supervisor only	-0.579	0.706	-0.82	0.412	-1.963	0.804	
Schooltype:							
High secondary (base)	-	-	-	-	-	-	
Comprehensive secondary	0.306	0.504	0.61	0.544	-0.682	1.295	
Practical secondary	-15.566	1.205	-12.92	0	-17.928	-13.204	***
Constant	-1.072	2.152	-0.50	0.618	-5.289	3.145	

Mean dependent var	0.037	SD dependent var	0.188
Pseudo r-squared	0.097	Number of obs	923
Chi-square	513.943	Prob > chi2	0.000
Akaike crit. (AIC)	283.499	Bayesian crit. (BIC)	331.840
*** p<.01, ** p<.05, * p<.1			

3.2. Item Nonresponse

Figure 2 shows the predicted item nonresponse given the three supervision scenarios, based on a negative binomial regression model. The maximum value would be 150. This is the highest possible number of unanswered questions for observations which are not considered to be drop-outs. While there is no significant difference between mode (B) and one of the others, the modes with only one adult present differ significantly in terms of item nonresponse. Students that were supervised by teachers only had an average item-nonresponse of five, while it was less than two for students that were supervised by external supervisors only. This could be interpreted as evidence in favour of hypothesis 1b and 2a, teachers decrease data quality while external supervision is a clear benefit.

The coefficients presented in Table 4 indicate that, with a one-unit change in the predictor variable and the other predictors held constant, the logarithm of the expected counts of the response variable (number of unanswered questions) changes by the respective value. For instance, the estimator for the effect of web surveys versus paper questionnaire is -0.788. The exponential of that yields 0.455, which means that the expected count of unanswered questions is approximately 45.5% lower for respondents using a web survey compared to those using a paper questionnaire, ceteris paribus. While that seems like a very large effect, accounting for other variables in the model and the constant is important to not overestimate the scale of the effect. School types also significantly differ, comprehensive and practical secondary schools have higher item-nonresponse than high secondary schools.
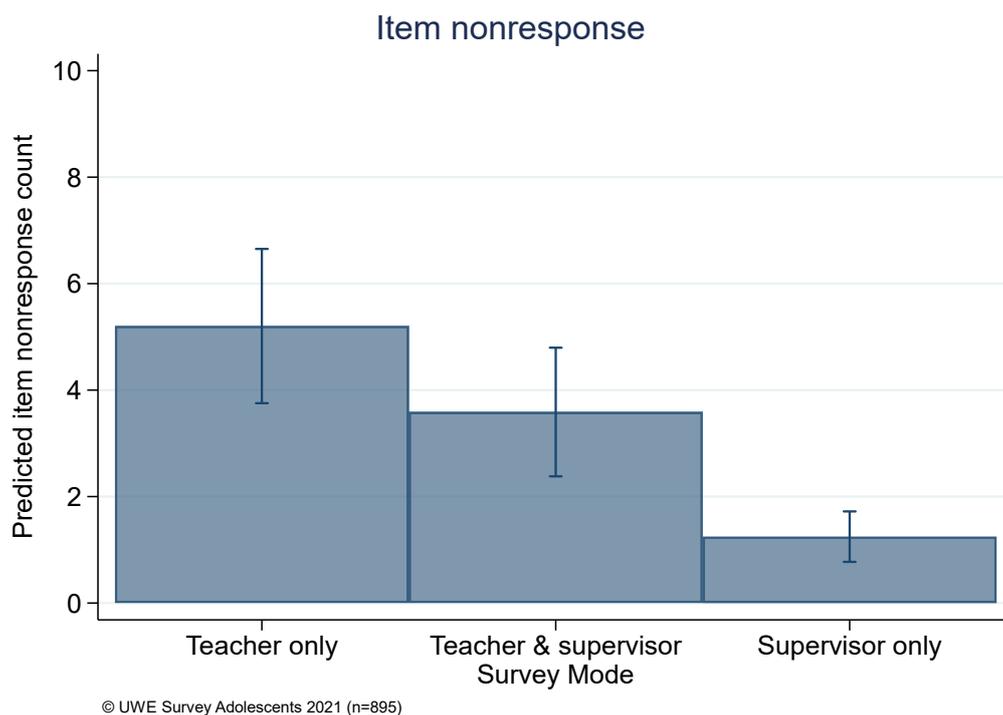
Figure 2. Predicted item nonresponse by mode of supervision (negative binomial regression, full figures in Table 4).

Table 4. Item nonresponse - negative binomial regression results.

	Coef.	St.Err.	z-value	p-value	[95% Conf	Interval]	Sig
Respondent characteristics:							
Gender	-0.165	0.171	-0.97	0.335	-0.499	0.17	
Age	0.02	0.057	0.35	0.729	-0.092	0.132	
Problems with reading German	-0.024	0.666	-0.04	0.971	-01.33	1.282	
Migration background	0.154	0.115	1.33	0.182	-0.072	0.38	
Interview characteristics:							
Web Survey	-0.788	0.378	-2.08	0.037	-1.529	-0.047	**
Teacher only (base)	-	-	-	-	-	-	
Teacher & supervisor	0.179	0.223	0.80	0.422	-0.259	0.617	
Supervisor only	-0.42	0.241	-1.74	0.081	-0.892	0.052	*
School type:							
High secondary (base)	-	-	-	-	-	-	
Comprehensive secondary	0.817	0.242	3.38	0.001	0.343	1.292	***
Practical secondary	1.86	0.388	4.79	0	1.099	2.62	***
Constant	0.41	0.888	0.46	0.645	-1.331	2.15	
lnalpha	0.475	0.101	-	-	0.278	0.672	

Mean dependent var	4.068	SD dependent var	9.815
Pseudo r-squared	0.042	Number of obs	895
Chi-square	169.814	Prob > chi2	0.000
Akaike crit. (AIC)	4158.635	Bayesian crit. (BIC)	4211.400
*** p<.01, ** p<.05, * p<.1			

3.3. Interview Duration

The survival curves reported in Figure 3 show the predicted share of online questionnaires that are still open at the time points given by the x-axis, based on a Cox regression. The dashed line represents interviews supervised by teachers only, the thin red line stands for the sessions with teachers and additional support by external supervisors. Both trajectories are hardly distinguishable. They differ from the third, thick green line. Students taking the survey at home, virtually supervised by external staff took more time on average to complete the survey. After 30 minutes, around 80% of the former were finished, while in the latter group less than 60% were done.
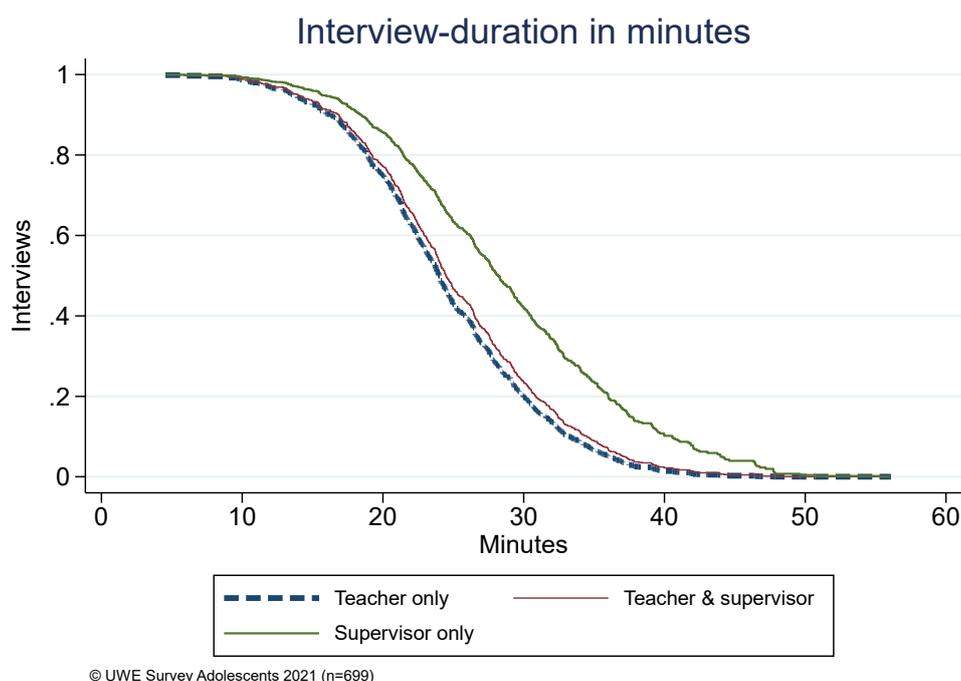

© UWE Survey Adolescents 2021 (n=699)

Figure 3. Predicted interview duration in minutes by mode of supervision (Cox regression, full figures in Table 5).

If the hazard ratio for an exemplary predictor is 1.2, it means that, on average, the hazard (risk) of the event occurring is 20% higher for each one-unit increase in the predictor variable. Assuming linear effects of the explanatory variables, each year of age means an increase in hazard of 20%, older students finished the survey faster. Interview duration is the only indicator that differs significantly by age. Conversely, if the hazard ratio were 0.56 (supervisor only mode), it would suggest that, on average, the hazard of the event is 44% lower for each one-unit decrease in the predictor variable. Interestingly, literacy did not have an effect. Schools types differ, with the high secondary schools hosting shorter sessions on average than the other two types.

Table 5. Interview duration - Cox regression results.

	Hazard ratio	Robust St.Err.	z-value	p-value	[95% Conf	Interval]	Sig
Respondent characteristics:							
Gender: Female	0.99	0.075	-0.14	0.892	0.853	1.149	
Age	1.206	0.049	4.60	0.000	1.114	1.307	***
Problems with reading German	0.863	0.273	-0.47	0.64	0.464	1.603	
Migration background	0.948	0.072	-0.70	0.484	0.817	1.1	

Interview characteristics:

Teacher only (base)	-	-	-	-	-	-	
Teacher & supervisor	0.921	0.123	-0.61	0.54	0.709	1.198	
Supervisor only	0.561	0.125	-2.60	0.009	0.362	0.867	***
School type:							
High secondary (base)	-	-	-	-	-	-	
Comprehensive secondary	0.654	0.091	-3.06	0.002	0.498	0.858	***
Practical secondary	0.23	0.073	-4.62	0.000	0.123	0.429	***

Mean dependent var	25.654	SD dependent var	7.935
Pseudo r-squared	0.013	Number of obs	699
Chi-square	62.718	Prob > chi2	0.000
Akaike crit. (AIC)	7683.672	Bayesian crit. (BIC)	7720.069

*** p<.01, ** p<.05, * p<.1

3.4. Straightlining

Figure 4 illustrates the predicted probability of manifesting suspected satisficing response behavior—discerned by the presence of a straight line in at least one out of six item batteries. On average, questionnaires completed in schools were more likely to contain straight lines in the grids. The difference is only significant between students solely supervised by teachers compared to students surveyed at home. The probabilities are relatively high for the former group with over 20%. The full figures of the analysis of straightlining feature two significant effects. Besides the mode of supervision, gender seems to make a difference in response behaviour: girls were less likely to resort to straightlining. In all other predictors, no effects can be reported.
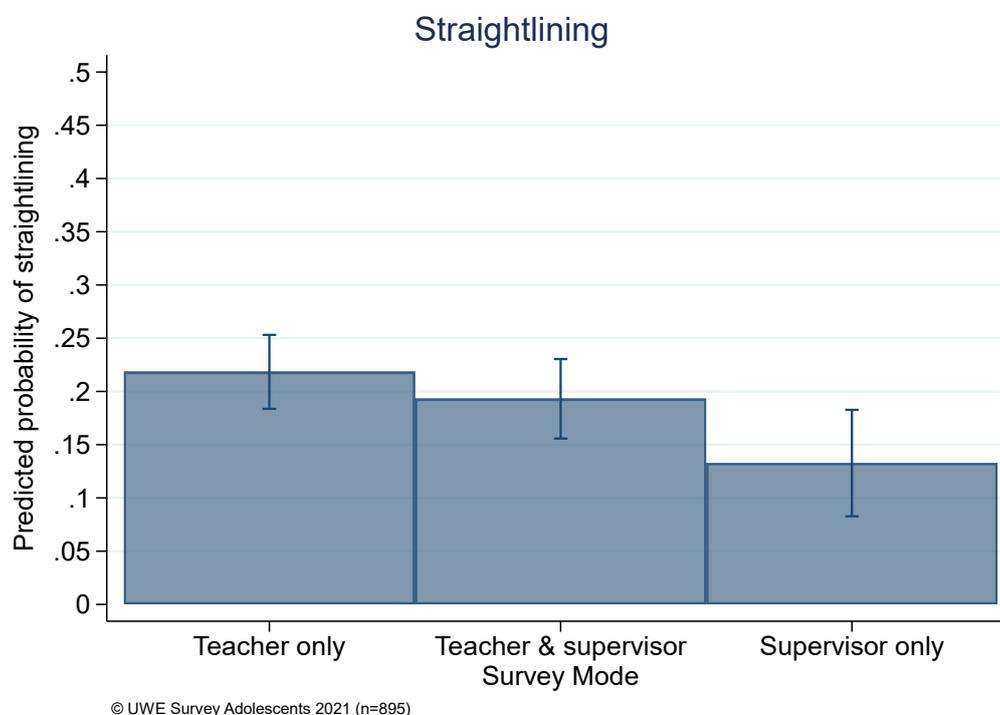

© UWE Survey Adolescents 2021 (n=895)

Figure 4. Predicted probability to have at least one occurrence of straightlining by mode of supervision (logistic regression, full figures in Table 6).

Table 6. Straightlining - logistic regression results.

	Coef.	St.Err.	z-value	p-value	[95% Conf	Interval]	Sig
Respondent characteristics:							
Gender: Female	-0.745	0.186	-4.01	0.000	-1.109	-0.381	***
Age	-0.084	0.066	-1.26	0.208	-0.214	0.046	
Problems with reading German	0.261	0.456	0.57	0.568	-0.633	1.154	
Migration background	-0.113	0.16	-0.71	0.478	-0.427	0.2	
Interview characteristics:							
Web survey	-0.234	0.292	-0.80	0.424	-0.806	0.339	
Teacher only (base)	-	-	-	-	-	-	
Teacher & supervisor	-0.283	0.209	-1.35	0.176	-0.693	0.127	
Supervisor only	-0.603	0.301	-2.01	0.045	-1.192	-0.014	**
Schooltype:							
High secondary (base)	-	-	-	-	-	-	
Comprehensive secondary	0.107	0.24	0.45	0.656	-0.363	0.577	
Practical secondary	-0.03	0.287	-0.10	0.917	-0.594	0.533	
Constant	0.356	0.947	0.38	0.707	-1.5	2.211	

Mean dependent var	0.198	SD dependent var	0.399
Pseudo r-squared	0.031	Number of obs	895
Chi-square	31.717	Prob > chi2	0.000
Akaike crit. (AIC)	882.369	Bayesian crit. (BIC)	930.337

*** p<.01, ** p<.05, * p<.1

4. Discussion

A successful recruitment process can be seen as a prerequisite for high data quality, and previous literature shows that schools can be quite efficient partners for this endeavour (Heath et al. 2009, Alibali & Nathan 2010, Bartlett et al. 2017, March et al. 2022, Ellonen et al. 2023). This study does not question this but advises to carefully assess potential effects of any involvement of school staff when it comes to data collection. Hatch et al. (2023) suggest working as closely with school officials and personnel as possible, a conclusion that is based on considerations about response rates in the first place. This claim is valid as the loss of a single school in the process can threaten representativity of the whole sample (Newransky et al. 2020, Ellonen et al. 2023). The main message of this publication is that besides recruitment, involvement of school staff in standardized survey research in classrooms should be limited to a minimum.

This rather challenging claim is based on evidence of previous research as well as the results presented above. While there is a reasonable claim that any supervision can increase completion rates and validity of survey data when surveying adolescents (March et al. 2020, Bidonde et al. 2023), supervision might not be unproblematic in any case. There are two key factors speaking against teachers supervising survey research: They can make lousy research assistants (Demkowicz et al. 2020; Rasberry et al. 2020; March et al. 2022) and they can increase social desirability or related response bias unintentionally (Strange et al. 2003; Duncan and Magnuson 2013; Atkeson et al 2014; Cops et al. 2016; Möhring & Schlütz 2019). The results of the analyses in this article allow the conclusion that data quality is lower when teachers are responsible for data collection with regards to item nonresponse and the prevalence of satisficing patterns.

In light of potential compromise of data quality, the literature suggests that the introduction of external supervisors, acting as neutral parties well-versed in survey conduct, may serve as a valuable countermeasure. External supervisors have been shown to increase motivation (Demkowicz et al. 2020; March et al. 2022) and contribute to a neutral atmosphere (Demkowicz et al. 2020), potentially enhancing the data collection process. While teachers are usually present in schools anyways,

(additional) external supervision is expensive (Walser and Killias 2012, Bartlett et al. 2017; March et al. 2022). Thus, the decision to involve them or not requires justification. This study tested two possible ways to ensure external supervision. Additional supervisors were present virtually in a video conference, while teachers were present in the classroom. This mode labelled "Mode B", had minimal effect on data quality. The efforts invested in establishing this mode were found to be disproportionately high in comparison to its impact on data quality. "Mode C" yielded considerably better results and did not involve teachers and classrooms in the first place. However, this study is clearly limited in that it cannot determine whether the enhanced data quality of this mode is a result of the supervisor or the setting.

Future research should ideally investigate the impact of external supervisors in classrooms in the absence of teachers and explore teachers' supervision of web surveys conducted in students' homes. Additionally, the data collection in UWE did not feature a completely unsupervised survey mode, which is a clear limitation. The "supervisor only" mode could be considered almost unsupervised, as there was no physical presence of any adult and no risk of responses being exposed to a third party. Mühlböck et al. (2017) examined differences between web surveys with interviewers present and modes without supervision and found none in terms of response behaviour but higher completion rates in the supervised groups. Both findings are contradicted by this study, as drop-out rates did not differ significantly, but satisficing response patterns were more prevalent in physically supervised modes compared to virtual supervision and are in line with Tourangeau and Smith (1996), who concluded that self-administration can yield better results indeed.

The study can contribute to the debate whether web-surveys or digital questionnaires work better in self-administered surveys, as opposed to paper forms. In addition to the evident advantage of eliminating the need for transcribing digitally delivered questionnaires, early research particularly supports this mode for its favourable impact on anonymity, reducing item nonresponse associated with social desirability (Tourangeau & Smith 1996; Hallfors et al. 2000). More recent studies indicate that the data quality of digital surveys is on par with traditional paper forms (Raat et al. 2017; Felderer et al. 2019), given that all formats and devices function as intended (Gummer and Roßmann 2015). Moreover, there is a notable inclination among younger respondents towards favouring digital delivery (Demkowicz et al. 2020). In contrast, the results presented here indicate a higher drop-out rate among respondents using the online questionnaire. There are two explanations for this which are backed by the protocols of the UWE study. Firstly, intermittent disruptions in the connection between digital devices and the survey server occurred. This was primarily attributed to the limited infrastructure in schools and respondents' devices occasionally lacking sufficient charge. Secondly, a paper form doesn't go away when you close it, posing a higher barrier to drop-out. However, when drop-outs were excluded from the analyses, item-nonresponse and prevalence of satisficing did not differ between the modes.

The quasi-experimental nature of this study could be considered a potential limitation, and it would be worthwhile to conduct further research using randomly assigned supervision modes. However, respondents did not self-select into supervision modes. The assignment was completely external and based on decisions of the school boards, depending on factors completely unrelated to respondents' competencies or motivation to fill out a questionnaire. Hence, the differences in item-nonresponse and response patterns are valid arguments for future research to not rely on teachers alone when conducting standardized survey research in classrooms.

5. Patents

References

Alibali, M. W.; Nathan, M. J. (2010). Conducting Research in Schools: A Practical Guide. *Journal of Cognition and Development*, 11(4), 397–407. doi:10.1080/15248372.2010.516417

Atkeson, L. R., Adams, A.N., Alvarez, M. (2014). Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys. *Political Analysis*, 22(3):304-320. doi: 10.1093/PAN/MPT049

Bartlett, R., Wright, T., Olarinde, T., Holmes, T., Beamon, E. R., & Wallace, D. (2017). Schools as sites for recruiting participants and implementing research. *Journal of community health nursing*, 34(2), 80-88.

Bidonde, J., Meneses-Echavez, J. F., Hafstad, E., Brunborg, G. S., & Bang, L. (2023). Methods, strategies, and incentives to increase response to mental health surveys among adolescents: a systematic review. *BMC Medical Research Methodology*, 23(1), 270.

Braekers, R., & Veraverbeke, N. (2005). Cox's regression model under partially informative censoring. *Communications in Statistics—Theory and Methods*, *34*(8), 1793-1811.

Brener, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., & Ross, J. G. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opinion Quarterly*, 70, 354–374.

Brown, B., & Larson, J. (2009). Peer relationships in adolescence. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 74–103). Wiley.

Cops, D., De Boeck, A., & Pleysier, S. (2016). School vs. mail surveys: disentangling selection and measurement effects in self-reported juvenile delinquency. *European Journal of Criminology*, 13, 92–110

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative research*, 14(5), 603-616.

Demkowicz, O., Ashworth, E., Mansfield, R., Stapley, E., Miles, H., Hayes, D., Burrell, D., Moore, A. & Deighton, J. (2020). Children and young people's experiences of completing mental health and wellbeing measures for research: Learning from two school-based pilot projects. *Child and Adolescent Psychiatry and Mental Health*, 14, 1-18.

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: moving from correlation to causation. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3), 377-386.

Ellonen, N., Pösö, T., Mielityinen, L., & Paavilainen, E. (2023). Using self-report surveys in schools to study violence in alternative care: A methodological approach. *Child Abuse Review*, e2814.

Epstein, D. (1998) '"Are you a girl or are you a teacher?" The 'least adult' role in research about gender and sexuality in a primary school', in G. Walford (ed.), *Doing Research About Education*. London: Falmer Press.

Felderer, B., Kirchner, A., Kreuter, F. (2019). The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias. *Journal of Official Statistics*, 35(1):93-115. doi: 10.2478/JOS-2019-0005

Gassman-Pines, A., Ananat, E. O., & Fitz-Henley, J. (2020). COVID-19 and parent-child psychological well-being. *Pediatrics*, 146(4).

Goh, E. C., & Binte Rafie, N. H. (2023). Using whatsApp video call to reach large survey sample of low-income children during covid-19: a mixed method post-hoc analysis. *International Journal of Social Research Methodology*, 1-16.

Gomes, H. S., Farrington, D. P., Maia, Â., & Krohn, M. D. (2019). Measurement bias in self-reports of offending: A systematic review of experiments. *Journal of experimental criminology*, 15, 313-339.

Gummer, T., & Roßmann, J. (2015). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*, 33(2), 217–234. https://doi.org/10.1177/0894439314533479

Gummer, T., Bach, R. L., Daikeler, J., & Eckman, S. (2021). The relationship between response probabilities and data quality in grid questions. *Survey Research Methods* (Vol. 15, No. 1, pp. 65-77). DEU.

Hallfors, D., Khatapoush, S., Kadushin, C., Watson, K., & Saxe, L. (2000). A comparison of paper vs computer-assisted self-interview for school alcohol, tobacco, and other drug surveys. *Evaluation and Program Planning*, 23(2), 149–155.

Hatch, L.M., Widnall, E.C., Albers, P.N., Hopkins, G.L., Kidger, J., de Vocht, F., Kaner, E., van Sluijs, E.M.F., Fairbrother, H., Jago, R. & Campbell, R.M. (2023). Conducting school-based health surveys with secondary schools in England: advice and recommendations from school staff, local authority professionals, and wider key stakeholders, a qualitative study. *BMC Medical Research Methodology* 23, 142 https://doi.org/10.1186/s12874-023-01957-x

Heath, S., Brooks, R., Cleaver, E., & Ireland, E. (2009). Researching young people's lives. Sage.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). Applied survey data analysis. CRC press.

Hennessey, A., Demkowicz, O., Pert, K., Mason, C., Bray, L., & Ashworth, E. (2022). Using creative approaches and facilitating remote online focus groups with children and young people: Reflections, recommendations and practical guidance. *International Journal of Qualitative Methods,* 21, 16094069221142454.

Kann, L., Brener, N. D., Warren, C. W., Collins, J. L., & Giovino, G. A. (2002). An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *Journal of Adolescent Health,* 31(4), 327–335.

Kivivuori, J., Salmi, V., & Walser, S. (2013). Supervision mode effects in computerized delinquency surveys at school: Finnish replication of a Swiss experiment. *Journal of Experimental Criminology*, 9, 91–107.

Knüttel, Katharina; Stefes, Till; Albrecht, Michaela; Schwabe, Katharina; Gaffron, Vanessa; Petermann, Sören (2021): Wie geht's Dir? Ungleiche Voraussetzungen für das subjektive Wohlbefinden von Kindern in Familie, Schule und Stadtteil. Bertelsmann Stiftung; ZEFIR. DOI: 10.11586/2021070.

Leiner, D. J. (2019). Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, 13(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Leiner, D. J. (2021). SoSci Survey (Version 3.2.24) [Computer software]. Available at https://www.soscisurvey.de

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198-1202.

Little, R. J. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, 87(420), 1227-1237.

Lucia, S., Herrmann, L., & Killias, M. (2007). How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology*, 3, 39–64.

March, A., Ashworth, E., Mason, C., Santos, J., Mansfield, R., Stapley, E., Deighton, J., Humphrey, N., Tait, N. & Hayes, D. (2022). 'Shall We Send a Panda?'A Practical Guide to Engaging Schools in Research: Learning from Large-Scale Mental Health Intervention Trials. *International Journal of Environmental Research and Public Health*, 19(6), 3367.

Möhring, W., & Schlütz, D. (2019). Das Interview als soziale Situation. In Die Befragung in der Medien-und Kommunikationswissenschaft (pp. 41-67). Springer VS, Wiesbaden.

Mühlböck, M., Steiber, N., Kittel, B. (2017) Less Supervision, More Satisficing? Comparing Completely Self-Administered Web-Surveys and Interviews Under Controlled Conditions. *Statistics, Politics, and Policy*, 8, 13-28. doi: 10.1515/SPP-2017-0005

Niu, L. (2020). A review of the application of logistic regression in educational research: common issues, implications, and suggestions. *Educational Review*, 72(1), 41-67.

Raat, H., Mangunkusumo, R. T., Landgraf, J. M., Kloek, G., & Brug, J. (2007). Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Quality of Life Research*, 16, 675-685.

Revilla, M., & Ochoa, C. (2015). What are the links in a web survey among response time, quality, and auto-evaluation of the efforts done?. *Social Science Computer Review*, 33(1), 97-114.

Schwabe, Katharina;   Albrecht, Michalea; Stefes, Till; Petermann, Sören (2021): Konzeption und Durchführung der UWE-Befragung 2019*. ZEFIR Materialien Band 17*. Bochum: Zentrum für interdisziplinäre Regionalforschung (ZEFIR).

Stefes, Till (2023). Umwelt, Wohlbefinden und Entwicklung von Kindern und Jugendlichen (UWE) Befragung 2021. GESIS, Köln. Datenfile Version 1.0.0, https://doi.org/10.7802/2613

Stefes, Till; Lemke; Annika; Gaffron, Vanessa; Knüttel, Katharina; Schuchardt, Jakob; Petermann, Sören (2023): Konzeption und Durchführung der UWE-Befragung 2021. *ZEFIR Materialien Band 22*. Bochum: Zentrum für interdisziplinäre Regionalforschung (ZEFIR).

Strange, V., Forest, S., Oakley, A., & Ripple Study Team. (2003). Using research questionnaires with young people in schools: the influence of the social context. *Int. J. Social Research Methodology*, 6(4), 337-346.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public opinion quarterly*, 60(2), 275-304.

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). The science of web surveys. Oxford University Press.

Türkan, S., & Özel, G. (2018). A Jackknifed estimators for the negative binomial regression model. *Communications in Statistics-Simulation and Computation*, 47(6), 1845-1865.

Weller, S. (2017). Using internet video calls in qualitative (longitudinal) interviews: Some implications for rapport. International *Journal of Social Research Methodology*, 20(6), 613-625.

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly*, 74(2), 223-243.

Wagner, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly*, 76(3), 555-575.

Walser, S., Killias, M. (2012). Who should supervise students during self-report interviews? A controlled experiment on response behavior in online questionnaires. *Journal of Experimental Criminology*, 8(1):17-28.

Ziegel, Eric, R.,. (1990). Survey Errors and Survey Costs. *Technometrics*, 32(4):466-467. https://doi.org/10.1080/00401706.1990.10484760